(51) International Patent Classification[7]: G06F 19/00

(21) International Application Number: PCT/EP02/00064

(22) International Filing Date: 7 January 2002 (07.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01100362.1   5 January 2001 (05.01.2001)   EP
01104070.6   20 February 2001 (20.02.2001)   EP

(71) Applicant *(for all designated States except US)*: LION BIOSCIENCE AG [DE/DE]; Im Neuenheimer Feld 515, 69120 Heidelberg (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: O'DONOGHUE, Sean [AU/DE]; Untere Neckarstraße 56, 69117 Heidelberg (DE). FRIES, Karsten [DE/DE]; Bergheimer Straße 101a,

69115 Heidelberg (DE). MEYER, Joachim [DE/DE]; Burnhofweg 7, 69126 Heidelberg (DE). SCHAFFER-HANS, Andrea [DE/DE]; Mozartstraße 5, 64342 Seeheim (DE).

(74) Agents: SCHOHE, Stephan et al.; Boehmer & Boehmer, Hollerallee 32, 28209 Bremen (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(54) Title: METHOD FOR ORGANIZING AND DEPICTING BIOLOGICAL ELEMENTS
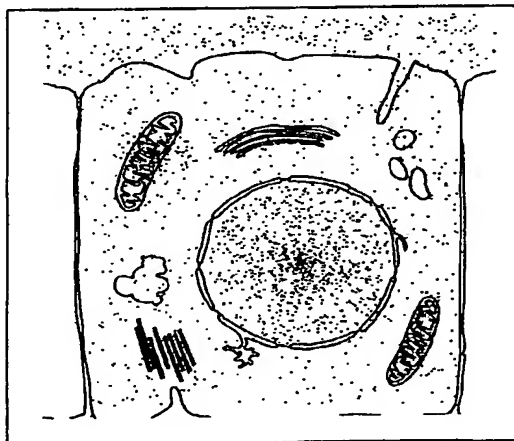
(57) Abstract: The invention relates to a method and apparatus for depicting one or more biological elements in a basic environment by means of a data processing system comprising the steps of obtaining one or more data sets relating to a biological element, determining at least one first feature element from said data sets, said feature element providing information on a relation between said biological element and said basic environment, obtaining data determining a graphical representation for depicting at least one of the biological elements corresponding to said one or more data sets determining a relations between the graphical representation of said basic environment and said graphical representation on the basis of said first feature element, providing means for effecting that in a graphical representation of said environment generated from said data said graphical representation of said biological element is depicted as located on said display of said basic environment element according to said relation determined on the basis of said first feature element. The invention also provides for a method and apparatus for handling three-dimensional representations of biological molecules, such as proteins and protein complexes.

Claims

1. Method for depicting one or more biological elements in a basic environment by means of a data processing system comprising the steps of:

    a) obtaining one or more data sets relating to a biological element,

    b) determining at least one first feature element from said data sets, said feature element providing information on a relation between said biological element and said basic environment,

    c) obtaining data determining a graphical representation for depicting at least one of the biological elements corresponding to said one or more data sets

    d) determining a relation between the graphical representation of said basic environment and said graphical representation on the basis of said first feature element,

    e) providing means for effecting that in a graphical representation of said environment generated from said data said graphical representation of said biological element is depicted as located on said display of said basic environment element according to said relation determined on the basis of said first feature element.

2. Method according to claim 1, comprising the further steps of

    -   receiving an input to select one or more databases,

    -   receiving further input to select one or more desired data sets from the one or more databases based on one or more selection criteria.

3. Method according to claim 1 or 2, wherein the biological elements are selected from the group comprising molecules, complexes of molecules, atoms, sub-atomic particles, SNPs/SAPs, exon/intron borders, annotations of active site residues and/or post-translational modification sites.

4. Method according to one of claims 1 to 3, wherein the display of said basic environment element is enhanced by locating at least 50 of said graphical representations of

said biological elements on said display of said basic environment element.

5. Method according to claim 4, wherein the display of said basic environment element is enhanced by locating at least 500 of said graphical representations of said biological elements on said display of said basic environment element.

6. Method according to claim 5, wherein the display of said basic environment element is enhanced by locating at least 5000 of said graphical representations of said biological elements on said display of said basic environment element.

7. Method according to one of claims 2 to 6, wherein one or more further feature elements are determined for said one or more data sets and the one or more further feature elements are extracted from the same database as the database from which the first feature element was determined.

8. Method according to claims 2 to 7, wherein one or more further feature elements are determined for said one or more data sets and the one or more further feature elements are extracted from one or more different databases as the database from which the first feature element was determined.

9. Method according to claim 7 or 8 comprising the steps of:
   - determining at least one second feature element for said at least one data set and,
   - further modifying the data representing the display of said basic environment element by using the information from said second feature element such that the graphical representation of said basic environment element is enhanced so as to place the at least one biological element into said basic graphical representation of said basic environment element.

10. Method according to claim 7 or 8 comprising the steps of:
   - determining at least one further feature element for said at least one data set and,
   - further modifying the data for displaying said at least one graphical representation of said biological element by using the information from said further fea-

ture element.

11.   Method according to claim 10 wherein a further feature element for said at least one data set pertaining to said at least one biological element contains information pertaining to one of the activities of said biological elements and the feature element is used to modify the graphical representation of said biological element by displaying the activity thereof.

12.   Method according to one of claims 2 to 11, wherein said one or more databases, are chosen from the group comprising databases that comprise data sets with information regarding biomolecules, organic molecules and/or inorganic molecules found in living organisms.

13.   Method according to claim 12, wherein the databases comprise data sets with information regarding genes and/or proteins and/or chemical compounds.

14.   Method according to any of the above claims wherein said basic environment element is a representation chosen from the group comprising an organism, one or more tissue types, an organ, a cell, an organelle, a sub-cellular compartment, a complex of molecules, a molecule, an atom and/or a sub atomic particle.

15.   Method according to one of claims 1 to 14, wherein said basic environment element is a linear polymer molecule or an ensemble of linear polymer molecules.

16.   Method according to claim 15, wherein said basic environment element comprises a three-dimensional protein structure and one or more biological elements are protein features, such as SNPs, SAPs, domain boundaries, exon/intron borders, annotations of active site residues and/or post-translational modifications.

17.   Method according to claim one of claims 1 to 16, wherein the representation is a two- or three-dimensional representation.

18.   Method according to claims 15 to 17, wherein a three dimensional representation of a linear molecule is displayed together with a one-dimensional representation of said

linear molecule.

19. Method according to claim 18, characterized in that said one dimensional representation is a character representation.

20. Method according to one of claims 18 or 19, wherein the graphical representation of said basic environment element is a three dimensional representation of a linear molecule, wherein one or more one dimensional representations of the sites of said linear molecule are provided on a graphical user interface, wherein one or more biological elements are represented in one or more of said one dimensional representations and wherein selecting a representation of a biological element in one of these one-dimensional representations results in said three-dimensional representation of said linear molecule to be enhanced by the selected biological element.

21. Method according to one of claims 1 to 20 , wherein the graphical representation of said biological element is different from the graphical representation of said environment element.

22. Method according to one of claims 16 to 20, wherein said linear molecule comprises a sequence of elements and differs from the sequence of a second molecule aligned thereto such that there is an optimum match of the elements at the aligned sites of said first and second molecules, wherein said biological element consists of one or more elements of said first molecule differing from the aligned sequence of said second molecule and said first feature element comprises information on the site or sites of said differing elements, wherein said biological element is represented at the sites indicated by said first feature in a manner different from the representation of said linear molecule.

23. Method according to one of claims 1 to 22, characterised in that a part of said basic environment element is selected on which whole or part of said biological element is to be located in said graphical representation.

24. Method according to claims 18 to 23, wherein a one-dimensional representation of a linear molecule is displayed; a range of elements of said molecule is selected in said

one-dimensional display, wherein said biological element is a part of said molecule and is depicted at the location of the three-dimensional representation of said molecule corresponding to the selected range in that one dimensional representation.

25.   Method according to any of the above claims comprising the step of determining the area that is depicted in said display of said basic environment element by selecting an area in a second display of said basic environment element.

26.   Method according to one of claims 15 to 22, characterised in that a range of said macromolecule displayed and/or selected and/or the element on which the cursor is currently positioned are displayed in said one-dimensional representation.

27.   Method according to one of claims 1 to 26, characterised by the steps of a method according to one of claims 28 to 93, 129 to 134 or 136 to 139.

28.   Method of visualising biological entities, especially linear macromolecules, by means of a data processing system, comprising the steps of
       -    providing a multidimensional graphical representation of one or more biological entities on a graphical user interface,
       -    interacting with said graphical user interface to select a point or feature of a portion of said graphical user interface containing said representation,
       -    determining a part of a biological entity to be selected that was not represented by a feature selected in said previous interaction step and selecting said part of said biological entity,
       -    performing an operation on the representation of the selected part of said entity.

29.   Method according to claim 28 wherein the representation of at least one biological entity comprises a plurality of individual graphical elements, said feature selected in said step of interacting with the graphical user interface is a graphical element of the representation of said biological entity and wherein a further part of said biological entity is determined to be selected that was not represented by said selected graphical element.

30.   Method according to claim 29, wherein said step of selecting a graphical element of

said representation implies the selection of the portion of said entity represented by said graphical element.

31. Method according to claim 29, wherein said step of selecting a graphical element of said representation implies the selection of only part of the portion of said entity represented by said graphical element.

32. Method according to one of claims 28 to 31, wherein said biological entity is a linear molecule and said method comprises the steps of:
    -   interacting with said multidimensional graphic representation to select a further graphical element of the multidimensional graphical representation of said molecule, thereby selecting the entire range of said molecule represented by the portion of said graphical representation from said first graphical element to said second, further graphical element,
    -   performing an operation on the representation of the selected range of said molecule.

33. Method according to one of claims 28 to 32, wherein said biological entity is a linear molecule and said step of determining a further part of said biological entity comprises the steps of:
    -   interacting with said multidimensional graphical representation to select a further graphical element of the graphical representation of said biological entity, thereby adding the portion of said biological entity represented by said graphical element to the selected part of said biological entity.

34. Method according to one of claims 28 to 33, comprising the step of providing data representing one or more biological entities, at least one of these entities being represented as consisting of basic biological units, especially residues of a sequence.

35. Method according to claim 34, wherein said data related to said biological entity comprise data assigning one or more basic biological units to one or more structural units of said biological entity, said structural units and said basic biological units forming a hierarchy in that each of said structural units comprises at least one basic biological unit and/or another structural unit,

said step of determining a part of a biological entity to be selected comprises the steps of:

- determining a level of hierarchy and
- selecting a structural unit at the determined level of hierarchy which comprises a previously selected basic biological unit or group of basic biological units of said biological entity.

36. Method according to claim 35, wherein in a first step one or more graphical elements of said representation are selected and in a further step all basic biological units and/or structural units are selected that comprise or correspond to basic biological units corresponding to any of said graphical elements of said representation selected in said first step.

37. Method according to claim 36, wherein one graphical element of said representation is selected by interacting with said graphical representation of said biological entity and that the structural unit at said level of hierarchy is selected which comprises all basic biological units of said biological entity corresponding to the selected element of said representation.

38. Method according to one of claims 35 to 37, wherein said level of hierarchy is the next higher level to that of the greatest structural unit comprised in the group of selected basic biological units.

39. Method according to one of claims 35 to 37 wherein said level of hierarchy is greater or equal to that of the greatest structural unit represented by a previously selected graphical element.

40. Method according to one of claims 35 to 37, wherein said level of hierarchy is in a predetermined relation to the level of hierarchy represented by a selected graphical element.

41. Method according to claim 40, wherein in a first step one or more basic biological units are selected, a graphical element is further selected as an anchor object and that said level of hierarchy is in a predetermined relation to the level of hierarchy repre-

sented by said anchor object.

42.   Method according to one of claims 35 to 41, wherein said selection of the level of hierarchy is effected by a keystroke and/or a mouse click.

43.   Method according to one of claims 28 to 32, wherein a further part of the biological entity to be selected is determined by a calculation having as input parameters parameters related to one or more selected points or features of the graphical user interface or one or more basic biological units and/or structural units previously selected.

44.   Method according to claim 43, wherein those basic biological units and/or structural units within a certain distance around a point or feature selected in said graphical user interface are determined and the basic biological units and/or structural units determined are selected.

45.   Method according to claim 43 or 44, wherein the part of said biological entity comprising the basic biological units and/or structural units determined by said calculation are displayed differently from the previous representation of said biological entity.

46.   Method according to one of claims 43 to 45, wherein the part of said biological entity comprising the basic biological units and/or structural units determined by said calculation are displayed differently from other parts of the representation of said biological entity.

47.   Method according to claim 43 to 46, wherein a representation of said basic biological units and/or structural units determined by said calculation is displayed and/or marked subsequent to said calculation.

48.   Method according to one of claims 28 to 47, wherein those basic biological units and/or structural units closest to a point or feature selected in a portion of a said graphical user interface containing said representation of said one or more biological entities are determined by said calculation.

49.   Method according to one of claims 28 to 48, wherein said interaction with said graphi-

cal user interface comprises a mouse click and/or a keystroke.

50. Method according to one of claims 28 to 49, wherein said selected part of the biological entity is represented differently from the rest of the biological entity.

51. Method according to claim 50, wherein said selected part is represented in a different representation style.

52. Method according to one of claims 28 to 51 wherein said representation is zoomed to the selected part of said biological entity.

53. Method according to claim 52, wherein said zoom is such that the selected part of said biological entity is fully displayed on the screen.

54. Method according to one of claims 52 or 53, wherein said zoom is continuous.

55. Method according to one of claims 28 to 54, wherein said operation comprises executing a link assigned to a selected part of said biological entity or a selected part of the representation thereof and displaying information provided by said link

56. Method according to one of claims 28 to 55, wherein at least one of said biological entities comprises a molecule and a binding surface of said molecule is calculated with regard to the selected point or feature.

57. Method according to claim 56, wherein said selected feature is the representation of a further molecule or part thereof.

58. Method of visualising biological entities, especially linear polymer molecules or ensembles of polymer molecules,
by means of a data processing system, comprising the steps of
- providing data representing at least one biological entity as consisting of basic biological units wherein said data related to said biological entity comprise data assigning one or more biological units to one or more structural units of said biological entity, said structural units and said biological units forming a hier-